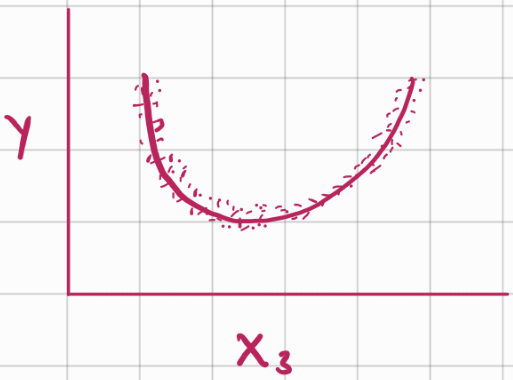
* If $Y = \beta_0 + \sum_{i=1}^{K} \beta_i X_i$



Possibly we have to perform polynomial regression or spline

We prefer linear models as the parameters are interpretable (have a physical meaning)

e.g : $wage = 5 + 10\, X_{age} + 20\, X_{age}^2$

- 5 → intercept or bias
- 10 → slope AED/yr given your age how much you earn
- $20\, X_{age}^2$ ↳ no physical meaning!

* Polynomial Model : Always take $\deg(P) \leq 4$

$$Y = \underbrace{\beta_0}_{\text{intercept}} + \underbrace{\beta_1 X_1 + \beta_2 X_2}_{\text{linear effect parameters}} + \underbrace{\beta_{11} X_1^2 + \beta_{22} X_2^2}_{\text{quadratic effect parameters}} + \underbrace{\beta_{12} X_1 X_2}_{\text{interaction effect parameter}} + \epsilon$$

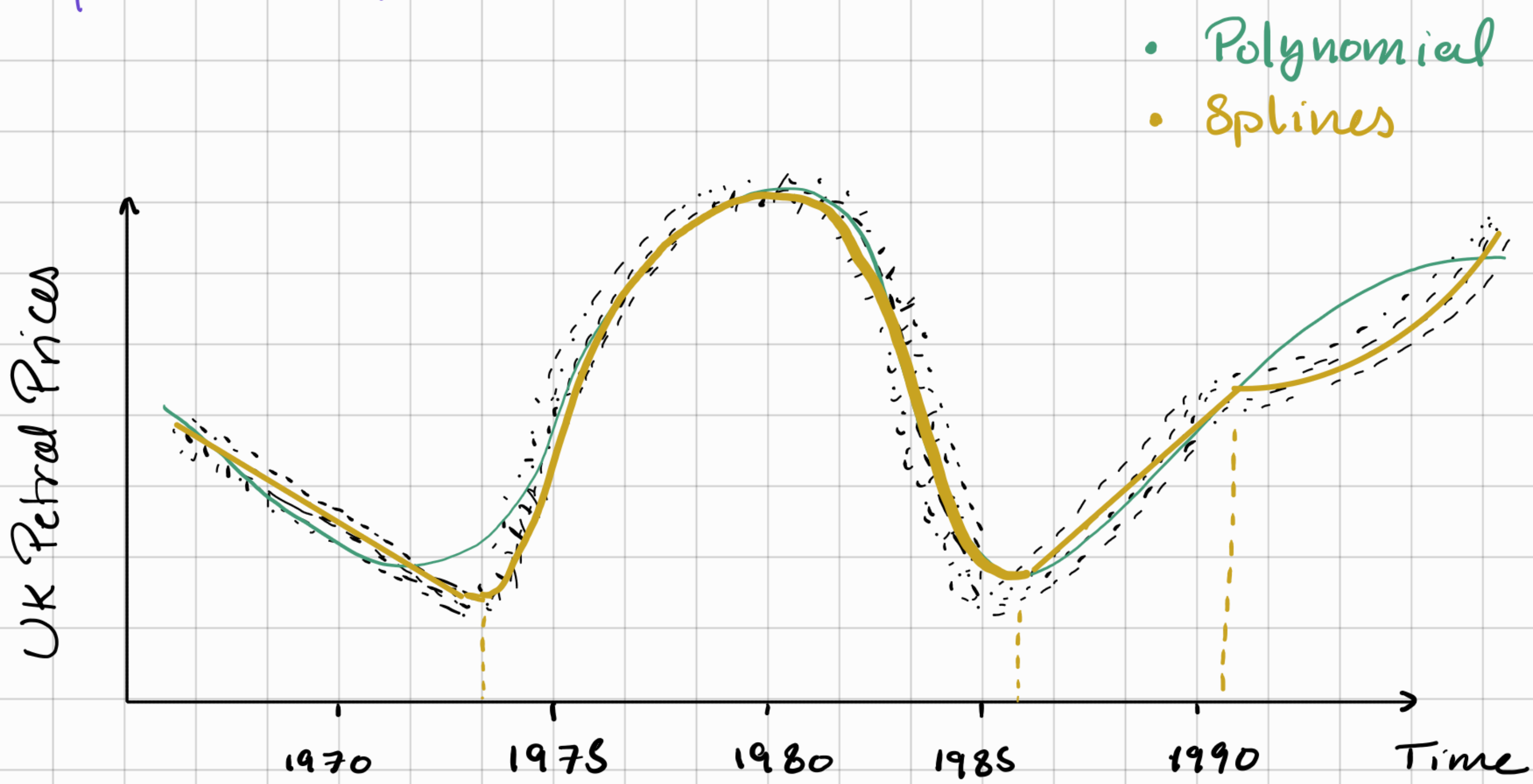
Response Surface (Regression Equation) :

$$E(Y \mid x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{11} x_1^2 + \hat{\beta}_{22} x_2^2 + \hat{\beta}_{12} x_1 x_2$$

$$E(Y|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$
$$= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{11} z_1 + \hat{\beta}_{22} z_2 + \hat{\beta}_{12} x_1 x_2$$

where $z_i = x_i^2$ so Y is linear with this change of basis and if Z is the design matrix then

$$\hat{\beta} = (Z^T Z)^{-1} Z^T Y$$



* Problems with each model :

1) Polynomial : Don't know at which degree we should stop

2) Splines : Deciding where the knots are

* Regression Spline (PW Polynomial) :

This is an extension of polynomial regression and piecewise constant regression approches. The joint points of the pieces are called knots.

* Linear Splines : $y_i = \beta_0 + \beta_1 b_1(x_i) + \dots + \beta_{K+1} b_{K+1}(x_i) + \varepsilon_i$

where $b_1(x_i) = x_i$

$$b_{k+1}(x_i) = h(x_i, \xi_k) = \begin{cases} x_i - \xi_k, & \xi_k < x_i \\ 0, & \text{ow} \end{cases}$$

where $k = 1, \dots, K$

* Generalised Cubic Spline with K knots :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X^2 + \beta_3 X^3 + \sum_{j=4}^{K+3} \beta_j b_j(x) + \varepsilon$$

Cross Validation helps in finding # of knots

The model is not smooth so future predictions may not be that good.

* Example : Chemical Process (Done in R)

Reaction Temperature (°C) X_1	Reaction Concentration (%) X_2	Conversion Y
200	15	43
250	15	78
200	25	69
250	25	73
189.65	20	48

260.35	20	76
225	12.93	65
225	27.07	74
225	20	76
225	20	79
225	20	83
225	20	81

Write the design matrix, find $\hat{\beta}$ and calculate SS_{Res}, SS_{Reg} and SS_T and do the anova table.